DOCUMENT RESUME

ED 365 118 FL 021 694

AUTHOR Coniam, David
TITLE Co-Text or No-Text? A Study of an Adapted Cloze Technique for Testing Comprehension/Summary Skills.
PUB DATE Sep 93
NOTE 11p.; For journal in which this paper appears, see FL 021 693.
PUB TYPE Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT Hong Kong Papers in Linguistics and Language Teaching; v16 p1-10 Sep 1993

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; *English (Second Language); Foreign Countries; Higher Education; *Language Tests; *Reading Comprehension; Testing; Test Validity; *Writing Skills
IDENTIFIERS Hong Kong; Summarization

ABSTRACT

This paper examines one section of the objective paper of the Hong Kong Examination Authority's (HKEA) Use of English Examination. One of the test types in this section of the examination involves candidates completing a cloze passage based on a text. It is the HKEA's contention that the exercise involves summarizing/paraphrasing skills. The findings of the current study, however, suggest that the validity of this test type is questionable since a substantial number of the blanks can be filled in without reference to the text. While it appears to be reliable and discriminates among candidates, as a test type, it would not appear to be achieving this by tapping the intended summarizing/paraphrasing skills. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

HONGKONG PAPERS IN LINGUISTICS AND LANGUAGE TEACHING 16 (1993) ISSN 1015-2059

Co-text or No Text? A Study of an Adapted Cloze Technique for Testing Comprehension/Summary Skills

ED 365 118

David Coniam[1]

This paper examines one section of the objective paper of the Hong Kong Examination Authority's (HKEA) Use of English examination. One of the test types in this section of the examination involves candidates completing a cloze passage based on a text. It is the HKEA's contention that the exercise involves summarizing/paraphrasing skills. The findings of the current study, however, suggest that the validity of this test type is questionable since a substantial number of the blanks can be filled in without reference to the text. While it appears to be reliable and discriminates among candidates, as a test type, it would not appear to be achieving this by tapping the intended summarizing/paraphrasing skills.

Introduction

This paper examines an adaptation of the cloze testing procedure - the summary cloze. This is one of the test types in the HKEA's Advanced Level Use of English Section C (Reading and Language Systems) examination (UEC). In this section of the examination, candidates complete a cloze passage based on a text. As this test type is in the examination paper, it is therefore widely practised in schools. It is the HKEA's claim that the exercise involves summarizing/paraphrasing skills. The opinion of the author, however, is that the validity of this test type is questionable, since a substantial amount of the blanks can be filled in without reference to the original text.

Background

Approximately 20,000 students take the UEC at the end of their final year in school, generally when they are about 18 years old (in Hong Kong this is called Form 7; it equates to Upper 6th Form in the U.K. and is one year above 12th grade in the U.S.A.). Candidates take the more specialised Advanced Level examinations two years after the more general Form 5 Hong Kong Certificate of Education (CE) Examinations. The candidature of UE is, then, understandably of a higher standard in terms of their English language proficiency, as well as being considerably more homogeneous in terms of ability in English than the comparable CE candidature.

The Use of English examination was totally revised from 1983 to 1988: so much so, in fact, that the when the 'new' examination was first administered in 1989, it bore virtually no resemblance at all to its predecessor. In re-designing the examination, the HKEA strove to include test types which tapped candidates' abilities at processing discourse, rather than test types which simply involved discrete sentence-level competence (which is more the focus of the Form 5 CE English language examination.) The UEC syllabus reads as follows:

> ...[the UEC examination] tests the extent to which the systems of the English language have been internalised by the candidates:

FL021694

[1]**David Coniam** has been involved in English language teaching and teacher education for almost 20 years. He has taught in the U.K., Spain, Iran, Saudi Arabia and Hong Kong. He is currently a lecturer in the Faculty of Education at the Chinese University of Hong Kong, where his main duties are teaching English Language Teaching Methodology. His main interests are computer analysis of language, language testing and methodology.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Yuen Yuen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1

2

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

- the lexicon

- the morphology

- the syntactic relationships within and among phrases and clauses, and

- the structural relationships among sentences within paragraphs and in discourse (1992 HKAL syllabuses, p.31)

It can be observed that the majority of the test types do focus on more than simply sentence-level proficiency.

The UEC examination is a 90-minute paper, divided into six sections of 'objective' type questions - three multiple-choice and three short-answer questions. In the design of this paper, it would have been all too easy to revert to discrete-item testing due to the short-answer nature of the items; nonetheless, the focus is on discourse-level skills. Of the six test types in UEC, only two - the multiple-choice cloze passage and the proof-reading exercise - have a focus which is more at sentence level than at discourse level, and it can be argued that even these two do involve skills of text processing across sentence boundaries. (See Johnson (1981) for a discussion of the extent to which MC cloze appears to be as sentence-based a test as individual discrete-point MC items.)

Focus of the current study

The cloze technique was devised by Taylor in the 1950s as a test of native-speaker reading ability. Since then it has undergone many modifications and extensions. From the original every nth deletion principle, we have seen variable deletion rates, deletions according to certain grammatical structures, and multiple-choice cloze, amongst others. More recently, Deyes (1984) puts forward a suggestion for a 'discourse cloze', where items are selected for deletion in such a way that the discourse context - whether cohesive ties, or lexical cohesion need to be taken into account. Lewkowicz (1991a, 1991b) puts forward a proposal whereby a type of paraphrasing cloze (somewhat similar to the HKEA summary cloze) is used to sample listening abilities.

The current study is an examination of the summary cloze. This is an adaptation of the cloze technique whereby the blanks are filled on the basis of a text the candidate reads. Reference to this type of exercise was originally made by Pollitt and Hutchinson (1987) with respect to testing L1 reading ability in U.K. secondary schools. Pollitt and Hutchinson were also advisers in the development of the International English Language Testing System examination (IELTS), which includes an adaptation of the summary cloze technique as one of its testing types for sampling reading comprehension. The credit for inclusion of the summary cloze in the UEC should go principally to Rex King of the HKEA, who had substantial involvement in the development of the new Use of English examination[1].

The idea of the summary cloze - as briefly outlined above - is that it tests comprehension in an indirect way in that candidates are required to paraphrase certain elements of a text that they have just read.

The attractions of the summary cloze technique on an objective paper such as the UEC paper are several:

- the paper needs to be able to be marked as simply and objectively as possible. The one-word blanks of the summary cloze are therefore reasonably straightforward in this regard.

- the HKEA did not want an examination which was totally multiple-choice (even though this would result in greater ease of marking). The fact that the summary cloze is not multiple-choice consequently improves the UEC's face validity.

- the summary cloze - as a test of reading - appears to tap discourse level skills. This again improves the UEC's face validity in light of the fact that it is not a discrete-point item test type.

- the summary cloze is a new test type in Hong Kong. While this is not a tremendous advantage in itself, it does mean that in the lower level forms in schools, it will not have been 'practised to death' as is the case with other test types. To a certain extent then, initially, there was some novelty value for teachers and students.

However, as a test type, the summary cloze does have its disadvantages. Through personal communication with teachers in secondary schools, the matter has been raised with this writer as to how the test type is different from the standard cloze passage as a test type. It has been suggested that the reading load demanded of candidates does not reflect the demands - or the rewards - of the task[2], in that candidates can complete certain blanks without reference to the text. The rubric in the summary cloze instructs candidates to:

> 'Read Version 2 [the cloze] and then fill in the blanks in Version 2 in such a way that the meaning of Version 1 [the original] is preserved.' (UEC 1989, p. 8)

There are a number of instances, however, where candidates can 'cheat'; where by applying their general reading skills, and by using their general knowledge - either of the world or their proficiency in English - they can supply an appropriate word, which is acceptable in terms of the marking scheme. For example, in the cloze passage under analysis, the first blank appears in the following context:

> "The 1st Battalion Coldstream Guards have now been in Hong Kong for nine months. As part of their training they recently staged an (1) _____ in the Sek Kong area."

The clues which the context offers for item 1 are, syntactically, that it is a noun, and one that begins with a vowel; and semantically that it is something that soldiers do as part of their training. The connection with a word such as 'exercise' or 'operation' should then not be too great a leap for a Form 7 student who is reasonably proficient in English. The facility value of this item on the 1989 examination was 72% - substantially higher than the overall summary-cloze part mean of 41%.

The passage continues:

> "Operation (2)_____ _____ involved 600 Guards in a mock action against 400 (3)_____ (4) _____, who acted as the (5) _____."

The facility value of item 5 is again high at 72%; the context again giving candidates considerable syntactic and semantic clues towards the word 'enemy'.

The focus of the current study is, therefore, the extent to which the summary cloze is tapping co-textual or extra-textual knowledge. The hypothesis is that the 'without-group' subjects who complete the summary cloze without the original text will achieve comparable results to the 'with-group' subjects who complete the summary cloze with the original text.

It is expected that mean and standard deviations of both groups will be quite similar, and that item analyses of both the with-group's summary cloze and the without-group's will be comparable with the 1989 whole group analysis. That is, the items that on the 1989 examination discriminated well, or

items which had high or low facility values, will demonstrate similar patterns with both groups.

Method

Two different test batteries were assembled from the 1989 UEC, each consisting of two parts. One consisted of the 20-item multiple-choice cloze passage, and the 21-item summary cloze passage together with the original text. The second comprised the multiple-choice cloze passage, and the 21-item summary cloze passage alone - with no text. The multiple-choice cloze passage was included in both tests to provide an indicator of both groups' abilities vis-a-vis each other compared with the 1989 whole group, as well as to provide a reliability anchor against which the two groups' performances could be measured on the summary clozes.

Two comparable groups in terms of ability level were then assembled. Some of these were drawn from Form 7 classes in Hong Kong secondary schools, as well as a small number of students from certain of the Chinese University's Faculty of Education Diploma in Education classes, the majority of whom were recent graduates. In each case, the class was divided arbitrarily into half, with approximately half (N = 152) randomly assigned the without-text test, and the other half (N = 145) the with-text test. As the timing of the six sections of the UEC examination is 90 minutes, it was decided that giving the groups 30 minutes to complete the two sections of their respective tests would approximate the live test time conditions. Care was taken that none of the subjects had taken or seen the paper before: for this reason the 1989 paper was selected for use.

The 1989 whole group figures for the multiple-choice cloze passage, the summary cloze, and the whole paper are presented in table 1:

Table 1: 1989 Whole Group figures

	MC cloze	Summary cloze	Whole UEC paper
No of items	20	21	99
Mean	67%	41%	55%
SD	15.6%	21.4%	13.0%

The HKEA considers 50-55% as the optimal mean for the various papers of its English language examinations. Table 1 above shows that while the overall UEC paper was on target here, the MC cloze was somewhat easier, but the summary cloze rather more difficult than the 'ideal'.

The results were then analysed under the following parameters:

- mean and standard deviations of both groups: to see how these compared with the 1989 whole group,
- Pearson product-moment correlations for both test groups between the multiple-choice cloze and the summary cloze: also for comparison with the 1989 whole group figures,
- item analyses of the two test groups,
- two-sample t-tests for the two groups, i.e. comparing group means to see if they differ

significantly.

Results

The mean and standard deviations of both groups are presented in table 2:

Table 2: Mean and standard deviations

	With-text Group		W/out-text group		1989 Whole group	
	MCC	SC	MCC	SC	MCC	SC
Mean	62%	38%	66%	26%	67%	41%
SD	18.6%	18.7%	17.5%	15.4%	15.6%	21.4%

[MCC = multiple-choice cloze; SC = summary cloze]

As can be seen from table 2 above, the multiple-choice means for the two groups are very close to the 1989 whole group figures: this demonstrates that in terms of English language ability, the results of the current sample are reasonably comparable with the Use of English whole group.

Table 3: Inter-subtest correlations

	With-text group		Without text group
	MCC		MCC
SC	.58 ($p < .001$)	SC	.44 ($p < .001$)

As table 3 illustrates, the subtests for the two groups both correlate significantly. In terms of student ability on the subtests, it can therefore be assumed that the more able subjects are performing well - in terms of their rank position on the test - on both summary cloze tests[3].

After applying the correction for attenuation to control for measurement error variance, the resulting corrected correlations were .77 for the with-test group and .61 for the without-test group. The shared variances of 59% for the with- and 37% for the without-test groups suggest that there is a considerable degree of overlap in what the two subtests are tapping. Given that the MC cloze arguably taps general proficiency rather than reading skills[4], a similar argument might therefore also be put forward regarding summary cloze. The focus of the current paper is, however, to investigate the extent to which summary cloze is a test of summarizing/paraphrasing skills rather than to speculate on the exact nature of the constructs that summary cloze is actually tapping.

Table 4: T-test significance levels

	Pooled t value	Variance DF	Estimate 2-tail prob.
MCC	1.77	295	.078
SC	-6.11	295	.000

As table 4 demonstrates, at the 5% level the t-test reveals no significant differences between the two groups on the MC cloze; it can be reasonably assumed therefore that the two groups are similar in terms of ability. Significance has emerged, however, between the two groups on the summary cloze; this will be discussed later.

Item analyses were run on the summary cloze test for the two groups. The results are presented in table 5.

Table 5: Item analyses of summary cloze items

Item	89WG	W/Text	Wo/Text	Item	89WG	W/Text	Wo/Text
1	58%	53%	58%	12	14%	12%	6%
2	72%	71%	* 1%	13	25%	16%	11%
3	25%	27%	* 5%	14	13%	10%	26%
4	21%	38%	*11%	15	36%	25%	18%
5	72%	63%	48%	16	45%	26%	12%
6	12%	8%	28%	17	43%	25%	11%
7	31%	34%	40%	18	46%	42%	35%
8	10%	9%	20%	19	56%	49%	63%
9	49%	41%	40%	20	68%	59%	38%
10	33%	33%	*15%	21	69%	59%	38%
11	68%	53%	* 1%				

* $X2(2)$, $p < .05$

Chi-square values were calculated for the 21 items to ascertain whether the observed facilities in the without-text group were significantly different from the 'expected' facilities of the other two groups. Item facilities significant at the 5% level ($df=2$) were identified. It can be seen that slightly over half the item facilities of the without-text group are comparable with the 1989 whole group's, and with the with-text group's.

The easiest item for the 1989 examination and the with-text group was item 2:

Operation _____ _____ involved 600 Guards ....
(VILLAGE FACTORY)
[Facilities: 1989 WG - 72%; With-text - 72%; Without-text - 1%]

This was an interesting item, in that it was the item which the with-text group and the 1989 whole group both found easiest; by contrast it was the most difficult item for the without-text group, as the answer was essentially a direct 'lift' from the text. Understandably, virtually no candidate in the without-text group was able to get the correct answer.

Item 8 was the most difficult for the 1989 whole group, and second most difficult for the with-text group:

....find hidden stocks of _____.
(WEAPONS)
[Facilities: 1989 WG - 10%; With-text - 9%; Without-text - 20%]

On this item, ironically, the without-text group did better than the with-text group and the 1989 whole group, possibly because the subjects had to infer from their general knowledge. The answer here was 'weapons'; since the text had distractors such as 'guns' and 'ammunition' (which were incorrect in the context), many of the subjects with the text scored no marks. Not having the text in this situation proved, in fact, to be an advantage!

An examination of the ten items from the without-text group identified as substantially below the other two groups in terms of facility values, reveals that they principally involve direct 'lifts' from the passage. It would appear that with items such as these the text is necessary since it is here that subjects have to employ their reading skills in order to locate the correct information, rather than inferring or guessing the answer from their general knowledge and/or using their general language proficiency.

In order to further test the assumption that the text was not necessary for the summary cloze, a 'linguistically naive' native-speaker informant was asked to complete the summary cloze, again without the passage. Her score was 13 out of 21, i.e 67%. Of the eight items that she got wrong, four (items 2, 11, 16 and 17) were asterisked as significantly different in table 5, and all appeared to involve direct 'lifts' from the text.

The native-speaker informant was then asked to complete the summary clozes without use of the texts for all the years that the UEC examination has been run. Her results were as follows:

Table 6: Native speaker informant scores without text

	1989	1990	1991
NS informant	67%	62%	48%
WG score (with text)	41%	52%	28%

As table 6 demonstrates, the text would not appear to be necessary in order to obtain a score of around 50%. The native-speaker's score has remained fairly constant, and would appear to correlate

with the actual difficulty level of the text. (The subject herself commented in fact that the 1991 summary cloze appeared much more difficult than the other two. This is supported by the statistics which show that the 1991 UEC candidature also found the summary cloze considerably more difficult than in previous years.)

While it can be argued that native speakers cannot but draw on general knowledge and reading skills in their interpretation of a text, if a non-native speaker subject can get 26% on a test such as the summary cloze without reading the passage, then there is a mismatch between the aims and the outcomes of the test.

Discussion

The null hypothesis that the text accompanying the summary cloze test makes no difference has, on the face of it, to be rejected. The t-test indicated significance, so it is not possible to say that the text simply does not make a difference: it does. The results need, however, to be viewed in the context of what is demanded of candidates on the test. In terms of ability, both groups in the current study are reasonably equivalent to the whole UE population. The direct comparison to be made would therefore appear to be that a group with the text scores 41% (8½ out of 21), and one without 26% (5½ out of 21). Although this 15% differential may look significant in terms of the test itself, in the context of the whole UEC paper, it equates to the with-text group scoring 3 more marks overall on the UEC paper through having invested 5-10 minutes reading the text and matching the questions to it. The questions to be asked here then are:

(1) To what extent it is fair that candidates spend 5-10 minutes processing a text, an investment of time which may not pay off in commensurate benefits in terms of added score?

- As the examination requires candidates to complete approximately 100 items in 90 minutes, if they are only going to be 3 marks better off, might they not be better advised to spend that time concentrating on other sections of the examination?

(2) Is it possible that the summary cloze passage under review was badly set?

- The fact that a native speaker consistently achieved approximately 50% without the text would appear to suggest that the 'problem' or whatever is occurring with the summary cloze in this study is not a weakness in design of this particular test but that it is endemic to the summary cloze test type in general. The papers at the HKEA are, however, set under the auspices of a rigorous committee which attempts to maintain parity across years, as well as scrutinizing the validity of the material and test types they are involved in setting[5].

(3) To what extent is the summary cloze actually tapping summarizing/paraphrasing skills?

- The answer here has to be a rather guarded one: if the without-text group's raw score is 26%, the with-text group's is 41%, and a native-speaker's is 67%, the answer has to be that some element of summary completion is being tested but only to a limited extent. The shared variance of 37% for the without-test group on the two subtests suggest that there may be some general factor - or factors - which both subtests are tapping. Without further investigation, however, this is a matter that cannot be speculated further upon at present.

Conclusions

A cynical recommendation for teachers, of course, would be that they should advise their students to maximise their time: this would entail them answering as much as they can of the summary cloze as an initial step without reading the passage. Then they should continue with the other sections of the examination, and only return to the summary cloze as the last section they attempt.

Should it be possible for the summary cloze to be better set? In theory, it would be desirable that every word required of candidates is a content word or a paraphrase from the passage alone - one that cannot be deduced from candidates' general knowledge or general proficiency. Given the fact that similar results emerged from a native-speaker informant completing the three summary clozes produced to date without the text, it would appear that the problems relating to this test type are not likely to be the result of poor setting.

The main argument that has been advanced is that the summary cloze as a test type for the testing of reading comprehension is low in construct validity. Any recommendations to the HKEA would have to be made in the light of how that body perceives the function of the summary cloze. If it forms part of the examination as an alternative yet reliable way of assessing general proficiency levels (but not specifically reading or summarizing/paraphrasing skills) in English, then the summary cloze is a viable - and valid - test type. If the HKEA takes the point of view that the summary cloze forms part of the UE examination to fulfill a specific purpose - i.e. what it currently purports to do - then its role in the examination needs to be re-examined: indeed, we might ask whether it should remain in the examination at all.

Notes

1. The name 'summary cloze' originated with King, although it is a term that Pollitt does not like (personal communication). He claims the test type was designed as a test of reading comprehension rather than simply as a variety of cloze test. He prefers the term 'Summary completion'.

2. Each blank in the summary cloze is worth one mark. This is the case for every item on the whole UEC, with the exception of the multiple-choice reading comprehension, where each question is awarded slightly more than one mark.

3. Such detailed test-type correlations are not released by the HKEA, only whole paper correlations; for 1989, these were in the region of .5 to .6.

4. Johnson (1981) describes an experiment where two sets of subjects were given an MC cloze - one group with the whole cloze, and the experimental group with the cloze as a set of scrambled discrete-point items. Performance was comparable across both groups, indicating that an MC cloze is not a text-based test: it is essentially a series of discrete-point testing items which tap general language proficiency. Johnson concludes:

> (...this cloze passage [...]intended as a measure of reading comprehension is not in any important sense a different measure from a set of equivalent discrete point items. (1981: 42)

5. The committee which actually sets/moderates the UEC has been together for a number of years and knows the test types they are involved in setting and moderating extremely well. (Personal communication)

References

Deyes, T. (1984). Towards an authentic 'discourse cloze'. *Applied Linguistics, 5(2),* 128-137.

Hong Kong Examination Authority. (1992). *Advanced Level Examination Regulations and Syllabuses.*

University of Cambridge Local Examination Syndicate. (1990). *International English Language Testing System*. Cambridge: University of Cambridge Local Examination Syndicate.

Johnson, R.K. (1981). The use of cloze procedure in the Form III English Language Secondary School Scaling Test. *Hongkong Working Papers In Linguistics and Language Teaching 4*, 28-47.

Lewkowicz, Jo. (1991a). Testing listening comprehension: a new approach? *Hongkong Papers In Linguistics and Language Teaching 14*, 25-31.

Lewkowicz, Jo. (1991b). Testing listening comprehension: can test writers be innovative? Paper presented at November 1991 Japan Association of Language Teachers' Conference.

Pollitt, A.B. and C.J. Hutchinson. (1987). *The English language skills profile: 14-16*. MacMillan Educational: Basingstoke, UK.

Taylor, W.C. (1956). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly 30*, 415-33.